DOCUMENT RESUME

ED 431 003 TM 029 829

AUTHOR Phelps, Richard P.
TITLE Education Establishment Bias? A Look at the National Research Council's Critique of Test Utility Studies.
PUB DATE 1999-05-00
NOTE 14p.; Originally appeared in: "Industrial Organizational Psychologist"; v36 n4 Apr 1999.
PUB TYPE Reports - Evaluative (142)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Aptitude Tests; *Bias; Employers; *Evaluation Problems; *Job Applicants; Occupational Tests; Research Methodology; Screening Tests; *Test Use
IDENTIFIERS *General Aptitude Test Battery; *National Research Council

ABSTRACT

It has been proposed that the federal government's General Aptitude Test Battery (GATB) might be used to provide private employers with the same kind of useful information the federal government and college counselors receive about job applicants. The U.S. Department of Labor asked the National Research Council (NRC) to review the issues related to use of the GATB for screening all job applicants and providing employers with the job applicant's GATB scores. The NRC formed a committee to review the issue, but the committee did not include any industrial organizational psychologists who had studied the GATB. The Committee recommended against the use of the GATB for a number of reasons. These reasons are reviewed, and reasons the report development process and the report itself are flawed are discussed. Criticism of the NRC process centers on the unusual composition of the review committee, the argument the committee made about the meager evidence of benefits from using the GATB, disagreement with the committee's conclusion about the "zero-sum labor market," and some contradictory positions taken in the report. The NRC committee is charged with bias in its evaluation. (Contains 1 figure and 22 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 431 003

Education Establishment Bias? A Look at the National Research Council's Critique of Test Utility Studies

Richard P. Phelps

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Richard Phelps
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

TM029829

BEST COPY AVAILABLE

Education Establishment Bias? A Look at the National Research Council's Critique of Test Utility Studies

Richard P. Phelps

While there exist "gatekeeping" tests for high school students going to college in the United States (the Scholastic Assessment Test [SAT] and American College Test [ACT]) and for high school students entering the skilled trades (apprenticeships plus licensing exams), in most U.S. states, there is no test for the other high school graduates, other than low-level "minimum competency" exams. Those high school graduates in or entering the "real world" in something other than a skilled trade are either tested at the discretion of individual employers with whom they seek work, or they are not tested at all. Moreover, there exist no gatekeeping exams at all for college graduates.

Much research evidence, manifest here and in other applied psychology journals, supports the proposition that a fairly standard general achievement or aptitude test would offer much of the predictive power to employers that the ACT and SAT offer college admissions counselors, if only employers would use one (e.g., Bishop, 1988; Boudreau, 1988; Hunter & Hunter, 1984; Schmitt et al., 1984). In their own national surveys, college admissions officers assert more confidence in SAT and ACT scores than in high school grade point averages. (National Association of College Admission Counseling, 1996) They are not devious, ignorant, or autocratic people, these college admissions counselors. They are likely more liberal, tolerant, and worldly than most of us, and they read the arguments on both sides of the issue, but they know from experience and research that general achievement or ability test scores are better predictors of performance than are high school grade point averages.

Proposed expanded use of the General Aptitude Test Battery

For years, it was proposed that the federal government's General Aptitude Test Battery (GATB) might be used to provide private employers the same kind of useful information the federal government and college admissions counselors receive. It seems fairly appropriate. The GATB is used to screen and place job applicants in the enormous labor pool of the U.S. federal government (see Hartigan; Wigdor, 1989, chapters 1,2,3).

The federal government, of course, has had an interest in knowing how useful the GATB is. Hundreds of predictive validity studies have been performed on data sets incorporating GATB scores, usually correlated with one or more job performance measures—that is, supervisor ratings, output-per-time period, promotions, earnings increases, and so on. Most readers are probably at least generally familiar with utility analysis, Brogden's formula, which provided a base for its empirical study, and the now fairly predictable conclusions of hundreds of related studies -- scores on general ability tests are better predictors of job performance than any other single predictor.

Essentially, general achievement scores probably demonstrate how hard a potential employee worked in school and how high the school standards were. Grade point averages only tell an employer how well a potential employee performed in school relative to other students at her school, if that. Grade point averages are norm-referenced measures, normed at the school level. Given

the absence of common or enforced standards in U.S. schools and the attendant enormous variety in their quality, it is no wonder that general achievement test scores explain a large amount of variance in regressions of job performance on groups of predictor variables that include school grade point averages.

John Bishop estimated that using a generalized achievement or aptitude test for job selection would produce an *annual* benefit of $850 to $1,250 per worker. Using Bishop's assumptions for calculating the present value (baseline at age 18 and, thus, a 45-year working life, and a 5% real discount rate) I calculate a range of present values between about $16,000 and $23,000 per worker over their working lives. By comparison with a cost of less than $50 per worker for such tests, the benefits loom enormous. (Bishop, 1988, 1994).

Bishop is estimating "job matching" or "allocative efficiency" benefits. John E. Hunter claimed benefits on a similar scale for "job selection" or "predictive validity" (Hunter, 1983). In efficient job matching, all workers in a market get assigned to jobs such that aggregate output is maximized. In efficient job selection, only the best workers get hired in the first place. Hunter claimed a potential benefit to the U.S. economy of about $80 billion (in 1980 dollars) from a 1-year application of the GATB for job selection in the entire U.S. labor market. (Hartigan; Wigdor,. 1989, pp. 237_238) That benefit calculates to about $20,000 per worker.

In the late 1980s, the U.S. Department of Labor seriously considered promoting the use of the GATB throughout the U.S. Employment Service to screen all job applicants for any jobs, not just those in the federal government. The employment service would then provide employers seeking workers each job applicant's GATB scores. (Hartigan; Wigdor, 1989, pp. iii_x)

The National Research Council Committee

The Labor Department requested that the National Research Council (NRC) review the issue and advise it on how to proceed. The NRC formed a Committee with an unusual membership. None of the hundreds of industrial-organizational psychologists who had studied the issue of the practical use of the GATB in testing for employment were invited. Of the 13 members, none were full-time members of university psychology departments (1 was part time in a psychology department). Four members, including the vice-chair, were education school professors and one worked for a consulting firm full time on education issues. The others were a mix from government, industry, and academe.

This Committee wrote a report, *Fairness in Employment Testing: Validity, Generalization, Minority Issues, and the General Aptitude Test Battery.* The tone of the document is not particularly respectful of the rich tradition of erudite research in utility analysis by I-O psychologists. The committee: criticized the validity studies of the GATB in several ways, driving down the predictive validity coefficient through a variety of rationales. They conceded a coefficient of 0.22, half the level of the highest, unadjusted predictive validity claimed for the GATB (Hartigan; Wigdor, 1989, pp. 134_171). Cutting the estimates of John Hunter above in half, however, produces present values of about $10,000 per worker lifetime, still enormous by comparison with the meager cost of a standardized test.

Then,in their chapter addressing the economic claims made for the GATB, the Committee claimed flatly that there are no job selection benefits testing because the U.S. labor market is a zero-sum game. If one employer selects

better workers by using GATB scores, the Committee argued, employers will get the other workers and it's all a wash. All workers work somewhere in the economy.

Analyzing the National Research Council Report

Several aspects of the NRC report *Fairness in Employment Testing* struck me, in addition to its bitter tone: (a) the odd composition of the committee; (b) the odd, repeated insistence of the committee that there was only meager evidence for the benefits of testing, in the face of hundreds of studies; personnel psychology research demonstrating those benefits; (c) the theory of the zero-sum labor market; and (d) the logical contradiction in the report's primary assertions that: all jobs are unique so *general* ability tests will be invalid for each, but there is no such benefit as a "selection effect" because any worker's abilities will be equally useful anywhere they work, no matter what their training and no matter what the field of work.

The Odd Composition of the Committee: Part I

Last year, I telephoned Alexandra Wigdor, the NRC study director and a co-editor of the report, to ask why researchers in personnel psychology were unrepresented on a panel about personnel testing and education school professors were so well represented. She asserted that there was no deliberate effort to exclude personnel psychologists or include education professors. They had sought out the best researchers they could find. It would have been improper, she continued, to have John Hunter on the committee, for example, as the committee would be focusing on his work, and he could be presumed to be biased in favor of it.

The NRC may not have been concerned about having committee members who could easily be presumed to be biased against Hunter's work, however.

The co-chair of the committee, Lorrie Shepard of the University of Colorado's School of Education, had just 2 years before conducted a "cost-benefit" analysis of a new basic literacy test for teachers in Texas in which the analysis was, if not ideologically biased, then very poorly done. Shepard's analysis contained arbitrary inclusions or exclusions of benefits or costs (see Phelps, 1996; Shepard, 1987).

For example, she counted the dismissal of teachers found to be illiterate as a benefit, because students would then be taught by the literate teachers who replaced them. However, in the fine print, one discovers that she decided that "nonacademic" teachers shouldn't be counted in the benefit calculations. Which teachers were "nonacademic?"—kindergarten, music, art, ESL, industrial arts, business education, physical education teachers, and counselors. No matter that the citizens of Texas wanted those teachers to be literate; Shepard decided they didn't need to be. Shepard also miscalculated the value of time by counting the benefit of the dismissed teachers for only 1 year, even though they were dismissed for good and the benefits would string out years into the future.

Shepard also counted costs of teachers' time spent studying for the tests, but no benefit to that studying, as if the teachers learned nothing by studying. Indeed, while she alleged many costs, she counted only that one benefit, from replacing illiterate teachers. There are at least several others.

After this exercise in maximizing costs and minimizing benefits was complete,

Shepard declared that the teacher test cost the citizens of Texas $53 million. Just adjusting for the mistakes in her own calculations changes the net present value to a positive $333 million. That's without adding the benefits she never mentioned.

The economists Lewis Solmon and Cheryl Fagnano estimated two other major benefits ignored by Shepard: the long-term labor-market benefits resulting from students learning more from more able teachers; and the attraction to the teaching profession of more able applicants as a result of higher professional standards (Solmon; Fagnano, 1990). They estimated these benefits to be as large as a billion dollars in present value. In another study, the economist Ronald Ferguson found teachers' literacy test scores to be the strongest predictor of Texas' minority students' success in school, stronger than any background variable (Ferguson, 1991).

The Odd Composition of the Committee: Part II

Though the National Research Council committee convened to investigate utility in personnel testing, none of the academic personnel psychologists involved in that research were included among its members. By 1989, there were hundreds who had conducted test utility analyses. Alexandra Wigdor implied, then, that the best researchers available to evaluate this personnel psychology research just happened to be education professors.

Out of curiosity, I made some calculations with binomial probabilities of the odds of picking only education school professors at random from a large pool of test researchers. Let's assume that personnel testing experts are equally distributed across places where personnel psychologists or education school faculty work. That's a big IF, but I want to be conservative. There are about 1,000 college professors in the National Council for Measurement in Education and about 3,900 persons total in the wider-scope Measurement and Research Methodology division of the American Educational Research Association. Similarly, there are about 1,500 persons in the American Psychological Association's Evaluation and Measurement division and over 5,000 in the Society for Industrial-Organizational Psychology.

Depending upon whether one circumscribes the fields narrowly or broadly, personnel Measurement experts outnumber school testing experts by a ratio of about 5.7 to 4.3. What are the odds that four school testing experts are the best qualified? Using the standard binomial probability formula, I calculate odds of only 0.03. If personnel measurement experts happen to be more qualified to judge personnel testing issues, the odds drop below even 0.03.

Would the federal government hire microeconomists to evaluate macroeconomic problems? Would it hire inorganic chemists to study an issue in organic chemistry? Would it hire personnel psychologists to evaluate school curricula? Why did the federal government hire education professors and education consultants to evaluate personnel testing issues?... especially given that the United States boasts some of the world's most advanced research and dozens of the world's most respected researchers in personnel testing.

The Meager Evidence of Benefits Argument

Consider the following quotes from ***Fairness in Employment Testing***:

It is also important to remember that the most important assumptions of the Hunter-Schmidt models rest on a very slim empirical foundation....Hunter and

BEST COPY AVAILABLE

Schmidt's economy-wide models are based on simple assumptions for which the empirical evidence is slight (p. 245).

Some 'fragmentary' confirming evidence that supports this point of view can be found in Hunter et al. (1988)... We regard the Hunter and Schmidt assumption as plausible but note that there is very little evidence about the nature of the relationship of ability to output (p. 243).

There is no well-developed body of evidence from which to estimate the aggregate effects of better personnel selection...we have seen no empirical evidence that any of them provide an adequate basis for estimating the aggregate economic effects of implementing the VG-GATB on a nationwide basis (p. 247).

...primitive state of knowledge... (p. 248).

Was the NRC Committee correct about the paucity of research? From the 1960s on, hundreds of studies have been conducted by dozens of researchers in personnel psychology affirming positive net benefits to the use of general ability testing in employee hiring.

There are so many studies it becomes more efficient to count just the *meta analyses*. A 1988 meta analysis by John Boudreau, then at Cornell University covered 87 such studies (Boudreau, 1988). A 1984 meta analysis by Schmitt, Gooding, Noe, and Kirsch covered over 300 studies (Schmitt et al., 1984). A 1997 paper by Schmidt and Hunter presented the validity of 17 different selection procedures over 85 years (Schmidt; Hunter, 1997). Hunter and Hunter conducted a *meta analysis of 23 meta analyses* in 1984, summarizing thousands of validity studies (Hunter & Hunter, 1984).

The Zero-sum Labor Market Argument: Part I

The NRC Committee asserted that, contrary to the claims of personnel psychologists, there are no job selection benefits to testing; the U.S. labor market is a zero-sum game. If one employer becomes more efficient in selecting good workers by using job applicants' GATB scores in making selection decisions, the Committee argues, some other employer will end up with those less efficient workers and it's all a wash. All workers work somewhere in the economy (Hartigan & Wigdor, 1989, pp. 241_242).

The zero-sum labor market argument is erroneous, in my opinion. First, there are the unemployed, comprising about 5% of the labor force. The Committee cites the fact that the unemployment *rate* is fairly stable over time as evidence that the unemployed population is stable (Hartigan & Wigdor, 1989, 235_248. While the *rate* may vary only within a narrow band, the labor market churns people through the ranks of the unemployed and marginally employed over and over.

Using figures from the Bureau of Labor Statistics for the average duration of unemployment (16.6 weeks) and the average number unemployed in 1995 (7.4 million), I estimate the number of individual "spells" of unemployment for 1995 at 23.2 million. That totals to 17.5% of the labor force unemployed *at some time during the year*. (U.S.BLS, Tables 2, 31, 35).

Another 3.3% of the labor force in 1995 were "economic part-time" employees. That is, they wanted to work full time but could not find full-time employment. Add them to the 17.5% above for a proportion of the labor force close to 21%. (U.S.BLS, 1968_96).

Then, there remain "contingent workers," whose number is very difficult to estimate. Anne E. Polivka calculates estimates ranging from 2.7 million workers in jobs less than a year, who expect the jobs to last no longer than 1 year more, to 6 million workers who simply do not expect their jobs to last. If we subtract the subpopulation of persons classified as "independent contractors or self-employed" from her upper bound, for the reason that those people have chosen a necessarily contingent occupation, we calculate 5.3 million workers who believe their jobs are temporary and probably do not want them to be. That total comprises 4% of the labor force. (Polivka, 1996)

These three subpopulations above—unemployed at some time during the year, economic part-time, and contingent workers—are 25% of the labor force. That still, however, does not include the large number of workers employed outside their field of training, like philosophy Ph.D.s who work as computer programmers, college graduates in international affairs who work as secretaries, and so on. These workers have jobs that require a lower level degree for entry. These workers are "underemployed".

Finally, there remain an estimated 8.6% of the adult population out of the labor force who have quit looking for work out of discouragement for heir prospects.

The National Research Council assumed that if a worker didn't get selected for a job, she would get selected for a different job and that other job would be equivalent in the most important ways to the job she didn't get. That assumption is untenable. The person not selected for the first job could end up unemployed (p = .175), unwillingly working part time (p = .033), working in contingent employment (p = .04), underemployed (p = ?), working in a field outside their training, or out of the labor force entirely. This is a large group of adults.

The Zero-Sum Labor Market Argument: Part II

Let's pretend that two college students graduate at the same time from different colleges with degrees in organizational psychology and enter the job market as Worker A and Worker B. They have approximately the same grade point averages, but Worker A attended a college with higher standards followed courses of more rigor, studied more, and studied harder than Worker B. Thus, while both workers A and B accumulated human capital in the field of organizational psychology and in general abilities, Worker A accumulated more than did Worker B, a human capital surplus. This surplus is not detectable from the college transcripts, however, or letters of recommendations, or work experience, which are the same for both A and B. The surplus is detectable only through testing.

I have diagrammed a (very) simple labor market for these two workers and two employers, X and Y, in Figure 1. The diagram specifies various hiring scenarios: under poor or strong economic conditions; with one or both jobs being in or not in the field of the workers' training; and with both, one, or neither employer testing the workers.

In strong economic conditions, both employers have jobs available; in poor economic conditions, only one employer has a job available.

If only one employer tests, that employer will become aware of Worker A's human capital surplus and will want to hire A but will only have to offer a slightly higher salary than the other employer offers Worker B. This is because the other employer is ignorant of Worker A's surplus and so sees workers A and B as equally qualified. The employer knowledgeable of Worker A's surplus will, thus, capture Worker A's surplus in the form of higher quality work, without having to pay more than a nominal amount for it. If both employers test, and both are aware of Worker A's surplus, then Worker A can bid them against each other up to the point where the anticipated benefit of her surplus is fully incorporated in her salary offer. With full information, Worker A is compensated for her surplus. If an employer's job is in the graduates' field of study, they should be more willing to pay for Worker A's surplus because she has more need of it.

The 12 possible outcomes of these various permutations are explicit in Figure 1. I propose that 3 of these outcomes contain benefits that can be ascribed to job selection and allocation effects. Outcome 1 contains the job selection benefits and outcomes 8 and 9 contain the job allocation benefits.

For outcome 1: Employer X is the only one with a job available in a poor economy; she tests the two job applicants; and Worker A is hired after scoring higher on the test. Because Worker A must take whatever salary is offered, the employer gets to pocket Worker A's human capital surplus. Without the test, however, employer would have hired Worker A with only a .5 probability and, thus, only a .5 probability of capturing the surplus. The test increases Employer X's probability of putting Worker A's surplus to use from .5 to 1.0.

For outcome 8, employers X and Y both have jobs available, but Employer X's job is in the same field so she needs Worker A's surplus more than Employer Y does. In this case, both employers test, and Employer X hires Worker A at a salary somewhere above Worker B's, and shares Worker A's surplus with Worker A. If Employer X did not test, her probability of capturing part of Worker A's surplus would be only .5, while the probability that some of Worker A's surplus would be wasted (if Worker A worked at the job outside her field) would also be .5. Thus, by testing, employer A increases the probability of putting the human capital surplus to use from .5 to 1.0.

For outcome 9, only Employer X tests and becomes aware of Worker A's surplus. She hires Worker A for only a slightly higher salary than was offered Worker B. By testing, Employer X increases the probability of hiring Worker A (and putting her surplus to use) from .5 to 1.0.

Outcomes 1, 8, and 9 have much in common. Each increases the probability, through testing, of putting Worker A's human capital surplus to use rather than letting it be wasted. Productive assets are employed, rather than left unused. How, then, is outcome 1 an example of a "job selection" benefit, while outcomes 8 and 9 are examples of "job allocation" benefits?

Perhaps the best way to understand the difference between the two lies with considering Worker B, the one who gets less when the test reveals Worker A's surplus. In "job selection" outcome 1, Worker B cannot get a job, so her human capital accumulation is wasted. In the case of the "job allocation"-affected outcomes 8 and 9, Worker B ends up with a job, but it is not in the field for

which she trained. Worker B spent years at a university majoring in organizational psychology and now waits tables; the human capital accumulation from her college years is wasted. She didn't need to go to college to learn the job of a server. Indeed, she could have spent those years as a server and would have been better off financially. The National Research Council would say that one employer's loss is the other employer's gain. But, what does the restaurant gain from Worker B's college training? It doesn't.

The NRC Committee also attempted to diminish the purported economic benefits of allocative efficiency, or job-matching. They chopped down a Hunter and Schmidt estimate of the benefits of 1.6 to 4% of GNP to just 1%, under the assumption that not all employers would use tests like the GATB to select employees nor use the tests optimally. Yet, 1% of the GDP is still over $80 billion. (Hartigan & Wigdor, 1989; pp.243_246)

Capturing even just the per-worker proportion of that 1% of the GDP worth of potential benefits would be large, far larger than the meager cost of administering a test. In an economy of $8 trillion GDP and 125 million people in the labor force, 1% equals $640 per worker of potential benefits. That's not nothing.

Logical Contradiction of Homogenous Jobs and Unique Tests

The NRC Committee claimed no selection benefits to employment testing: Employment Service use of the VG-GATB will not improve the quality of the labor force as a whole. If employers using the Employment Service get better workers, employers not using the Employment Service will necessarily have a less competent labor force. One firm's gain is another firm's loss... The economy as a whole is very much like a single employer who must accept all workers. All workers must be employed (Hartigan & Wigdor, 1989, pp. 241_2).

Essentially, the NRC argued that skills measured by employment tests are equally useful in all jobs. That, of course, assumes that general intellectual aptitudes or abilities are equally valuable in all lines of work, and covary equally with all other relevant skills, say those used in brain surgery or street sweeping.

At the same time, in its chapter 8, "GATB Validities," the NRC Committee asserted that "Validities vary between jobs... GATB validities have a wide range of values over different jobs." In order for preemployment tests to be beneficial, they must be uniquely tailored to unique jobs (Hartigan & Wigdor, 1989, pp. 170_171).

The two assertions from chapters 8 and 12 are contradictory. The NRC Committee tries to have it both ways: Declaring the GATB to be invalid in predicting job performance because every job is unique and, at the same time, declaring selection effects moot because any worker not getting one job will get another and provide equal value to society.

Conclusion and Discussion

I have spoken with three persons intimately familiar with the activity of the National Research Council's Committee on the General Aptitude Test Battery. After considerable deliberation of the available evidence, I reach the following judgments.

BEST COPY AVAILABLE

One person claims that the Committee was deliberately set up to be a hostile committee. I think the odds are strong that that claim is correct.

Another person claims that the Committee considered only one personnel testing study from among hundreds in existence, yet made claims that implied they had considered all of them. I believe this assertion is also true.

The third claims that the Committee refused to consider some of the most basic and relevantevidence pertaining to personnel testing issues, such as: the ways in which the Hunter and Schmidt estimates of utility *underestimated* the benefits of testing; the true magnitude of the effect of range restriction on the utility estimates (for which the Committee refused to correct); the true value of average interrater reliability of ratings of .50 (they assumed .80, thus undercorrecting for criterion unreliability); and (pertaining to the NRC assertion that Hunter and Schmidt did not adjust their estimates for the time value of money, incremental validity, or what have you) the substantial research in personnel psychology that has explicitly considered all those issues (and found little difference in the direction or magnitude of the resulting utility estimates).

This is a serious charge, that those at the National Research Council responsible for the evaluation of testing issues were (and remain) biased. Yet, I believe it to be true, and I believe that any fair-minded person who looked at the evidence would agree.

The National Research Council is supposed to represent the pinnacle of objectivity, the "court of last resort" on controversial research issues. Alas, I believe, it represents neither on testing issues. It seems biased-biased in conformity with an "education establishment" perspective.

References

Bishop, J. (1988a). The economics of employment testing. *Working Paper* #88_14, School of Industrial and Labor Relations, Center for Advanced Human Resource Studies, Cornell University.

Bishop, J. H. (1998b). Employment testing and incentives to learn. *Journal of Vocational Behavior, 33*, 404_423.

Bishop, J. H. (1994a). Signaling the competencies of high school students to employers. *Working Paper* #94_18. CAHRS, ILR, Cornell University.

Bishop, J. H. (1994b). Schooling, learning and worker productivity. in Rita Asplund, Ed. *Human Capital Creation in an Economic Perspective*. Helsinki: Physica-Verlag.

Boudreau, J. W. (1988). Utility analysis for decisions in human resource management. *Working Paper* #88_21, School of Industrial and Labor Relations, Cornell University.

Farkus, S., Johnson J., and Duffett, A. (1997). *Different drummers: How teachers of teachers view public education*. New York: Public Agenda.

Ferguson, Ronald F. (1991). Paying for public education: new evidence on how and why money matters. *Harvard Journal on Legislation, 28*(2), 465_498.

Hartigan, J. A.,Wigdor, A. K. (1989). *Fairness in Employment testing: validity generalization, minority issues, and the general aptitude test battery.* Washington, DC: National Academy Press.

Hunter, J. E. (1983a). *The economic benefits of personnel selection using ability tests: a state of the art review including a detailed analysis of the dollar benefit of u.s. employment service placements and a critique of the low-cutoff method of test use,* Washington, DC: Department of Labor, Employment and Training Administration.

Hunter, J. E. (1983b). *Test validation for 12,000 jobs: an application of job classification and validity generalization analysis to the general aptitude test battery.* Washington, DC: U.S. Employment Service, Department of Labor.

Hunter, J. E., Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96,* (1).

Hunter, J. E., Schmidt, F. L. (1982). Fitting people to jobs: the impact of personnel selection on national productivity. In Dunnette, M. D. and Fleishman, E. A., Eds. *Human performance and productivity: Volume 1—human capability assessment.* Hillsdale, NJ: Lawrence Erlbaum Associates.

National Association for College Admission Counseling. (1996). Members assess 1996 recruitment cycle in eighth annual nacac admission trends survey. *News from National Association for College Admission Counseling,* October 28, 1996, 2,4.

Phelps, R. P. (1996). Test basher benefit-cost analysis. *Network News & Views, Educational Excellence Network,* 1_16. (http://www.edexcellence.net)

Phelps, R. P. (1998). The Demand for Standardized Student Testing. *Educational Measurement: Issues and Practice. 17,* (3). 5_23

Phelps, R. P. (1999). Why testing experts hate testing. Washington, DC: Thomas Fordham Foundation.

Polivka, A. E. (1996). A profile of contingent workers. *Monthly Labor Review,* Washington, DC: U.S. Department of Labor.

Schmidt, F. L., Hunter, J. E. (1997). The validity and utility of selection methods in personnel psychology: Practical and theoretical implication of 85 years of research findings. Unpublished manuscript.

Schmitt, N., Gooding, R. Z., Noe, R. D., Kirsch, M. (1984). Metaanalysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology, 37,* 407-422.

Shepard, L. A., Kreitzer, A. E., and Grau, M. E. (1987). A case study of the texas teacher test. *CSE Report No. 276,* CRESST/UCLA.

Solmon, L. C., Fagnano, C. L. (1990). Speculations on the benefits of large-scale teacher assessment programs: How 78 million dollars can be considered a mere pittance. *Journal of Education Finance, 16,* (1).

BEST COPY AVAILABLE

U.S. Department of Labor, Bureau of Labor Statistics (1997). *Household Data: Annual Averages*. Washington, DC: U.S. Department of Labor Bureau of Labor Statistics. Tables 2, 31, 35, and unpublished tabulations.

Biography

Richard P. Phelps is an education economist based in Paris who writes on issues of testing and international indicators. The author would like to thank Frank Schmidt, Scott Oppler, Deb Wetzel, Chris Sager, and John Hunter for their help and advice. The author retains all responsibility for errors

End Notes

1. For a deeper analysis on the net benefits of the SAT, see Phelps, R. P. (1999).

2. John Bishop has written much about how good high school students only get paid what they're worth after several years of having to prove themselves all over again in the workplace because there exits no good means of signaling their competence to employers at the outset. See Bishop (1994a)

3. See Phelps, 1996 for an explanation of several other errors in Shepard's analysis.

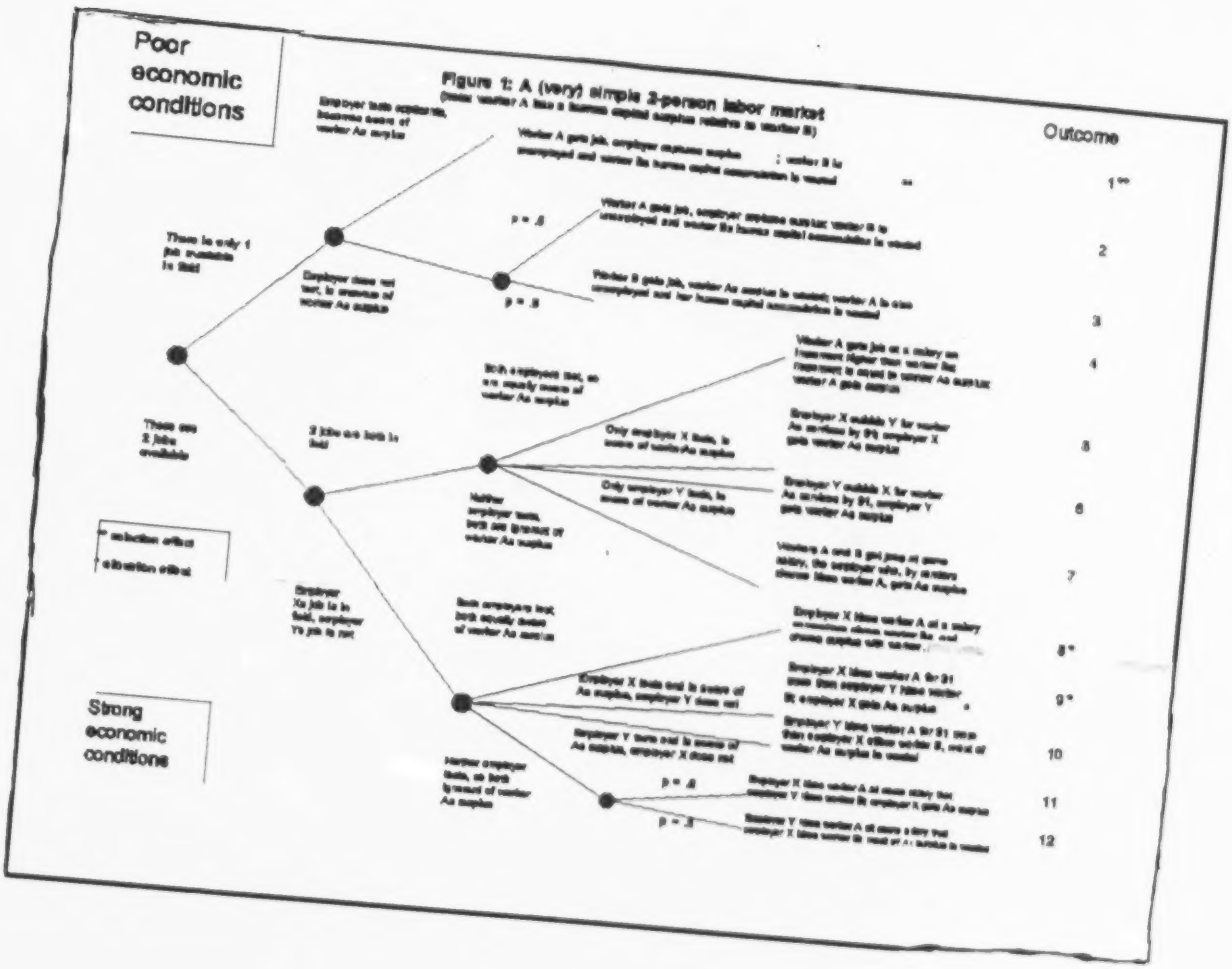
4. Only two members of the committee had any background in personnel psychology: one worked as an executive in a large corporation; the other worked in an administrative position at a university. Neither of them, however, was intimately familiar with the research on test utility, the studies of the GATB, and employee hiring. Several very well known personnel test utility researchers were included in the "Liason Group," but that group was little consulted and kept wholly unfamiliar with the secret deliberations of the committee.

5. Do education professors, in general, have policy preferences similar to the general public's, qualifying them to make policy decisions for the rest of us? Not on testing issues. See Phelps, 1999, pp. 1_2 and Conclusion, for a discussion of the Testing items in a 1997 Public Agenda poll of education professors.

6. At first thought, one might think that I am calculating the number of persons who are unemployed at some time during the year. While the estimate probably brings us close to that number, the estimate probably also subsumes a small number of spells that are shared by individuals. In other words, some persons may have more than one spell of unemployment in a year.

7. There is no average duration figure with which to calculate the number of persons who go through "economic part time" spells during the year. We have to settle for this lower-bound number for the number of workers who are at some time during the year forced to accept part-time employment when they would prefer full-time employment.

BEST COPY AVAILABLE


Poor economic conditions
Figure 1: A (very) simple 2-person labor market
Outcome
There is only 1 job available in field
There are 2 jobs available
2 jobs are both in field
Employer does not test, in excess of worker As supply
p = .5
p = .5
Both employers test, so are exactly aware of worker As surplus
Only employer X tests, in excess of worker As surplus
Only employer Y tests, in excess of worker As surplus
Neither employer tests, both are ignorant of worker As surplus
Employer Xs job is in field, employer Ys job is not
Both employers test, both equally aware of worker As surplus
Strong economic conditions
Neither employer tests, so both ignorant of worker As surplus
p = .5
p = .5
1**
2
3
4
5
6
7
8*
9*
10
11
12

BEST COPY AVAILABLE

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche, or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature: Richard P. Phelps	Printed Name/Position/Title: Richard P. Phelps	
Organization/Address: 1000 Thomas Jefs. St, NW ste 400 Wash, DC 20007	Telephone: +33.1.45.27.84.56	Fax: +33.1.45.24.90.98
	E-mail Address: sfurlott, cybercable.fr.	Date: 11 May, 1999

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:	The Industrial-Organizational Psychologist c/o Lee Hakel
Address:	SIOP Administrative Office P.O. Box 87 Bowling Green, OH 43402-0087 (tel.) 419-353-0032 (fax) 419-352-2645 (Email) lhakel@siop.bgsu.edu

V. WHERE TO SEND THIS FORM: